International Journal of Bio-Technology and Research (IJBTR)
ISSN (P):2249-6858; ISSN (E):2249-796X
Vol. 7, Issue 3, Jun 2017, 47-56

TRANS STELLAR
•Journal Publications •Research Consultancy

COMPARATIVE ANALYSIS AND INSILICO CHARACTERIZATION OF LEA, DREB AND OsDHODH1 GENES INVOLVED IN DROUGHT RESISTANCE OF *ORYZA SATIVA*

KAVALI RAGHAVENDRA[1], ABHINAV KUMAR[2] & AMIT KUMAR[3]

[1,2]*Noida International University, Noida, Uttar Pradesh, India*

[3]*Bio Axis DNA Research Centre Pvt Ltd, Hyderabad, Telangana, India*

ABSTRACT

Paddy being the most stable food crop of South India, has been facing a high demand for supply. There are several stress conditions faced by the paddy cultivators which may include salinity, drought, floods, harsh climatic changes, etc. Withstanding these parameters are the potential property of the Paddy variant, which is attributed to its genetic makeup. Some of the genes responsible for making the plant stress tolerant include LEA, DREB, OsDHODH1 etc. and are considered for the current study. The work is an extension of our previous study on extraction of these genes and sequencing. The sequences of these genes are analyzed both structurally and functionally, so as to implement the crop improvement strategies, if required. The three genes are also subjected to the conservation study of their genetic and protein sequences. The results of the comparative studies revealed that the three genes do not share any sequence similarity or conservation among their genes and protein sequences is absent. The protein network study conducted, failed to detect any common legends or proteins being interacted in their metabolic pathways. This study can conclude that at the sequence level of these three genes do not share any similarity or conservation in spite of their functional similarity in stabilizing the plant under stress conditions.

KEYWORDS: LEA, DREB, OsDHODH1, Drought Tolerance, Genetic Conservation, Insilico Analysis

Received: Mar 20, 2017; **Accepted**: Apr 14, 2017; **Published**: Jun 19, 2017; **Paper Id.**: IJBTRJUN20177

Original Article

INTRODUCTION

Oryza Sativa has been the most commonly used plant species for the laboratory investigations. This is the most important crop of south India and highest consumed food. In spite of its importance and high demand, there are certain parameters that are limiting the crop yield annually. It is desirable on the part of a molecular biologist and the agriculturist to act against this stress related parameter and overcome the shortage of the food supply to the population.

Apart from the number of Research protocols undertaken in the Molecular Biology, the implementation of Insilico strategies for management of high yield and development of quality products has made its platform. There are several *insilico* tools and software that enable the user to deal in detail with the genetic and proteomic data of the study sample. The computational protocols can add on to the laboratory research, which might be several times much detailed and clearer as compared to the molecular techniques.

The three genes involved in imparting the drought and stress tolerance of the Paddy crops include LEA, DREB and OsDHODH1. All the three genes are known for their similar effect of plant stabilization from the stress.

In view of this point the current work aimed to perform a comparative study at the genetic level of these three genes and their evolutionary relationship established which revealed a very diverse genetic makeup of these genes.

In the current study, an analysis on Sequence, Structure, Network and phylogeny of the three genes was performed and the conclusions were drawn based on their degree of similarity with respect to the plant resistance.

MATERIALS AND METHODS

Sequence Retrieval from NCBI

NCBI is a public library or database containing information related to the proteins, genes, SNP, Domains, Primers etc. It is one of three primary databases existing and is redundant in nature. NCBI's Gene and Protein Databases have been used to retrieve the sequences of all the three Genes and their respective proteins. The FASTA format of the sequence along with some important information like accession number, functional regions, active sites etc have been noted.

Conservation and Phylogenetic Study Using CLUSTAL W of EMBL

In order to detect the sequence conservation at both genetic and proteomic levels, all the three genes and respective protein sequences of the three genes have been compared using the CLUSTALW tool from EMBL. The conservation studies were followed by the evolutionary tree development. The tool identifies the degree of similarity, gaps and the conservation present among the three sequences. Based on the results of the multiple sequence alignment, the phylogenetic tree can be constructed.

Codon Plot

This tool is available at bioinformatics.org and is used to detect the exact frame from where the coding of the gene starts. This information is helpful in detecting the important codons that may be involved in binding site or SNP or any other function. The codon plot is also helpful for the detection of SNP effect in the protein translation. The triplet codon pattern of the three gene sequences is performed and the results are recorded. The codon plot is available at the SMS suit of Bioinformatics.org. The codon plot was performed for all the three gene sequences and their coding pattern has been recorded.

SMART Domain Analysis

SMART is a protein analysis tool that is used to identify the Domains i.e. the functional regions present in the user entered sequence. It detects the domains based on an inbuilt database of protein domains known. The user entered sequence is compared to all the sequences from the inbuilt database and the identical ones were recognized whose domain information is provided. The domain information includes the total number and types of domains, their length, location and function. Further the regions of low complexity are also highlighted in the results if any. SMART analysis was performed for all the three proteins and the results were recorded.

STRUCTURAL STUDIES OF THE PROTEIN

Primary Structure Analysis uses Protparam

Protparam is a tool from Expasy, which is used for the complete physic chemical characterization of the protein. This tool can be considered under primary structure prediction tools. The tool can predict the sequence based properties of the protein like the Length, Molecular Weight, Hydropathicity, Stability Index, Isoelectric Point and Half-life. These parameters intern help in the development of experimental protocol.

Secondary Structure Analysis uses SOPMA

For the identification of the internal structural conformations like loop, coil, turn, etc. Secondary structure has to be analyzed. SOPMA is one such tool used to analyse the conformational folds within the protein structure. Along with the structural conformations a graphical representation of the summary is also provided.

Tertiary Structure Prediction Using PHYRE

Phyre is a 3D structure prediction tool that runs on a BLAST Like algorithm that can compare the user entered sequence with the list of protein sequences stored in the PDB database and the result is displayed as the best PDB Id's that share maximum sequence similarity. The results here are the PDB ID'S of the structures that are already stored in a Protein data bank. One can visualize the structure by downloading this PDB from the protein data bank.

RASMOL Visualization of the PDB Structures

Rasmol is special visualization software that is required to view the 3D structure of the protein. This is a command line program that can be used to dictate the 3D structure of the protein. This software can be used to highlight the regions of interest, show the legend and water molecules present in the protein structure and also focus the ligands bound to the protein structure.

RESULTS AND DISCUSSIONS

Protein and Gene Sequence Retrieval

The protein sequences of all the three proteins have been retrieved and the sequence characters were analyzed. The length of the protein sequence of all the three genes LEA, DREB and OSDHODH1 are 200aa, 298aa and 414aa respectively. The sequences were further analyzed with respect to their function, structure, etc. In additional to the retrieval of protein sequence the gene sequences were also retrieved from the same data base. The gene sequences were 603bp, 897bp and 1245bp in length.

Comparison of the Gene and Protein Sequences

All the three genes and their corresponding protein sequences were compared wrt the sequence similarity using Clustal W programme, their evolutionary relation is measured using the phylogenetic tree constructed from Clustal W programme. The results are shown below.

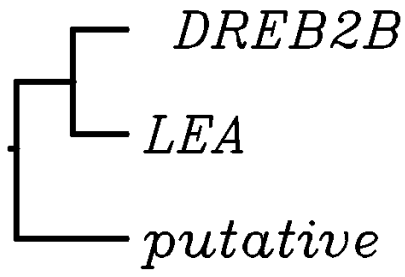

Figure 1: The Phylogenetic Tree of the Three Genes LEA, DREB and OsDHODH1

The above three indicates a closer relatedness of DREB with LEA which can be considered as Parallouges evolving partially were as the OSDHODH1 (Putative) shows a different evolutionary branch which indicates no similarity. The three sequences do not share any sequence level similarity either at the genetic level or at the protein stage. This can

be clearly identified in their CLUSTAL W Multiple sequence alignment.

Codon Plot

The triplet coding pattern of the three genes has been analyzed to detect the exact frame by which the translation can be performed insilico. The results of the same indicate that all the three genes are coding from its first base i,e the coding pattern starts from the beginning of the sequences. The results are shown below.

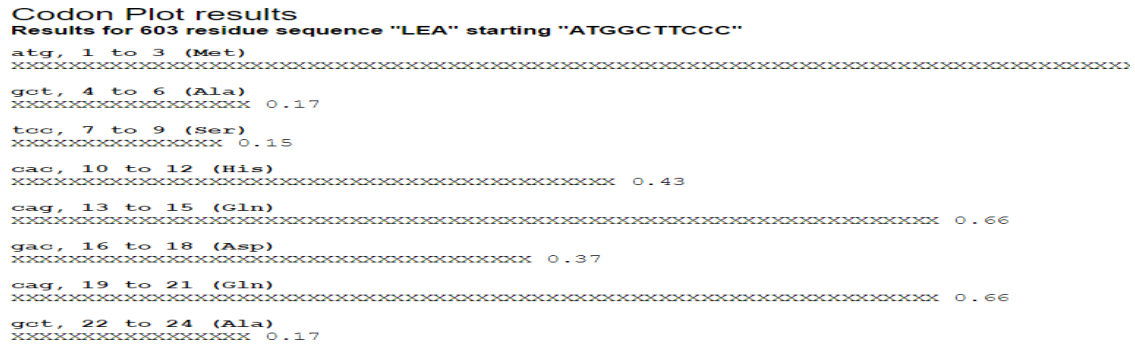

Figure 2a: Showing the Codon Plot Result for LEA

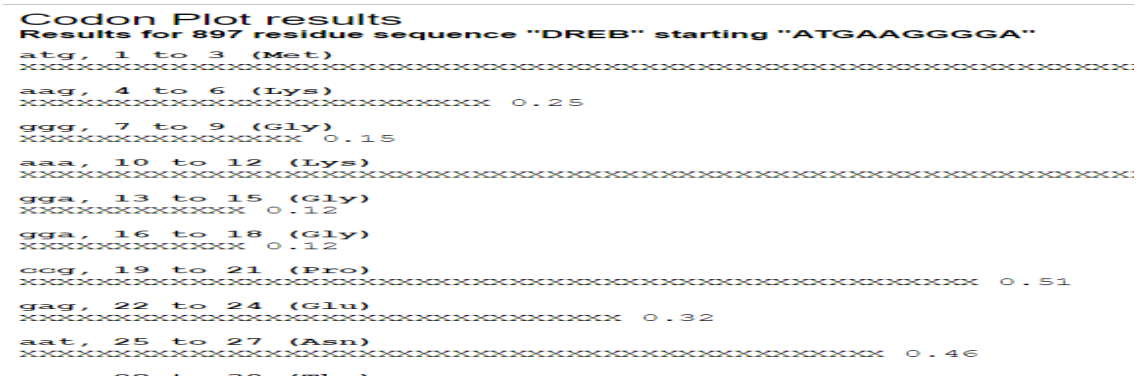

Figure 2b: Showing the Codon Plot Result for DREB

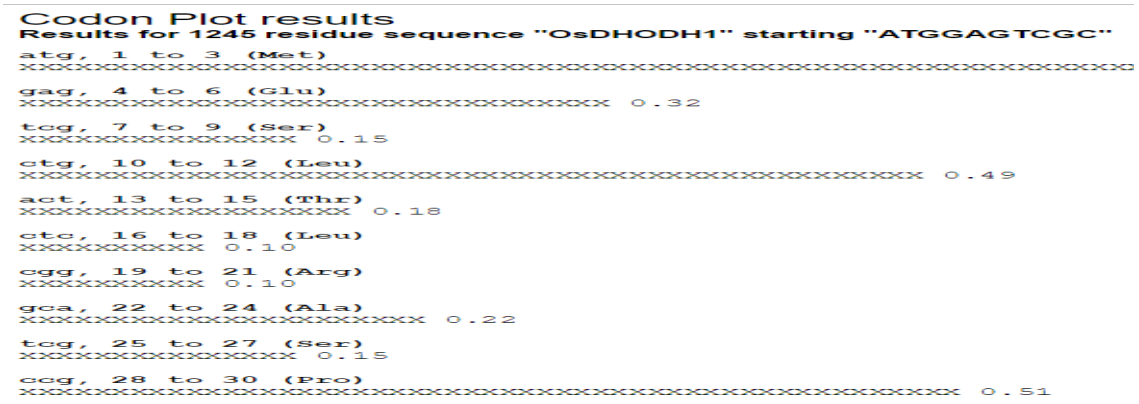

Figure 2c: Showing the Codon Plot Result for OsDHODH1

Domain Analysis Using SMART

The domain analysis of the three genes was performed to detect the functional units within the three protein sequences. Also a comparative study was performed to detect any common domain in the three.

Figure 3a: The domain Analysis of LEA

The above result shows that there is no individual domain detected in the LEA gene. However, two regions of Low complexity have been detected at 104bp and 186bp location.

Figure 3b: Domain Regions of DREB

As shown in the above figure, the only domain seen in DREB gene is AP2 which is located from 13 to 76 locations in the protein. This is the DNA binding domain present on the plant species. This is involved in transcription regulation process.

The domain analysis of OsDHODH shows that the data is totally not available pertaining to this gene.

Structural Analysis of Proteins

Protparam Analysis for the Three Proteins

In order to study the physicochemical characteristics of the three proteins Protparam has been employed. The tool details certain information like the length, stability, half-life, hydropathicity etc of the protein, which can be later used for the development of experimental protocol. The protparam results of the three proteins have been depicted below.

LEA Protein:

Number of amino acids: 200

Molecular weight: 20512.10

Theoretical pI: 5.88

Total number of negatively charged residues (Asp + Glu): 33
Total number of positively charged residues (Arg + Lys): 30

Extinction coefficients:

This protein does not contain any Trp residues. Experience shows that this could result in more than 10% error in the computed extinction coefficient.

Extinction coefficients are in units of M^{-1} cm^{-1}, at 280 nm measured in water.

Ext. coefficient 2980
Abs 0.1% (=1 g/l) 0.145

Estimated half-life:

The N-terminal of the sequence considered is M (Met).

The estimated half-life is: 30 hours (mammalian reticulocytes, in vitro).
>20 hours (yeast, in vivo).
>10 hours (Escherichia coli, in vivo).

Instability index:

The instability index (II) is computed to be 25.10
This classifies the protein as stable.

Aliphatic index: 38.50

Grand average of hydropathicity (GRAVY): -1.065

Figure 4a: Physicochemical Characteristics of LEA Protein

Inference

From the above results, it can be inferred that the LEA protein has a length of 200 amino acids with the molecular weight to be 20512.10Kd. The protein is basic in nature with more positive amino acids. It is a stable protein with instability index 25.10 and is hydrophilic in nature.

DREB PROTEIN:
Number of amino acids: 298

Molecular weight: 32158.39

Theoretical pI: 4.28

Total number of negatively charged residues (Asp + Glu): 52
Total number of positively charged residues (Arg + Lys): 23

Extinction coefficients:

Extinction coefficients are in units of M^{-1} cm^{-1}, at 280 nm measured in water.

Ext. coefficient 36690
Abs 0.1% (=1 g/l) 1.141, assuming all pairs of Cys residues form cystines

Ext. coefficient 36440
Abs 0.1% (=1 g/l) 1.133, assuming all Cys residues are reduced

Estimated half-life:

The N-terminal of the sequence considered is M (Met).

The estimated half-life is: 30 hours (mammalian reticulocytes, in vitro).
>20 hours (yeast, in vivo).
>10 hours (Escherichia coli, in vivo).

Instability index:

The instability index (II) is computed to be 51.22
This classifies the protein as unstable.

Aliphatic index: 64.93

Grand average of hydropathicity (GRAVY): -0.525

Figure 4b: Physicochemical Characteristics of DREB Protein

Inference

From the above result of protparam it can be concluded that DREB gene is 298aa in length with a corresponding molecular weight of 32158.39 kd. The isoelectric point of the protein was found to be 4.28 and the stability index was 51.22 indicating that the protein is unstable. It is a basic protein, which is polar in nature.

OsDHODH1 PROTEIN:

Number of amino acids: 414

Molecular weight: 45308.90

Theoretical pI: 6.29

Total number of negatively charged residues (Asp + Glu): 51
Total number of positively charged residues (Arg + Lys): 48

Extinction coefficients:

Extinction coefficients are in units of M^{-1} cm^{-1}, at 280 nm measured in water.

Ext. coefficient 45295
Abs 0.1% (=1 g/l) 1.000, assuming all pairs of Cys residues form cystines

Ext. coefficient 44920
Abs 0.1% (=1 g/l) 0.991, assuming all Cys residues are reduced

Estimated half-life:

The N-terminal of the sequence considered is M (Met).

The estimated half-life is: 30 hours (mammalian reticulocytes, in vitro).
>20 hours (yeast, in vivo).
>10 hours (Escherichia coli, in vivo).

Instability index:

The instability index (II) is computed to be 42.04
This classifies the protein as unstable.

Aliphatic index: 81.52
Grand average of hydropathicity (GRAVY): -0.260

Figure 4c: Physicochemical Characteristics of OsDHODH1 Protein

Inference

From the above result of protparam for OsDHODH1, it can be inferred that the protein is 414 amino acids in length with a corresponding molecular weight of 45308.90 kd. The protein is a basic protein with the instability index found to be 42 making the protein unstable. The hydropathicity of the protein was found to be -0.260 making it a polar protein in nature.

Secondary Structure Prediction of the Three Proteins using SOPMA

SOPMA tool details the secondary structural conformations of the protein. The tool has been employed to detect the structural conformations of the three proteins. The results are depicted in the below figures.

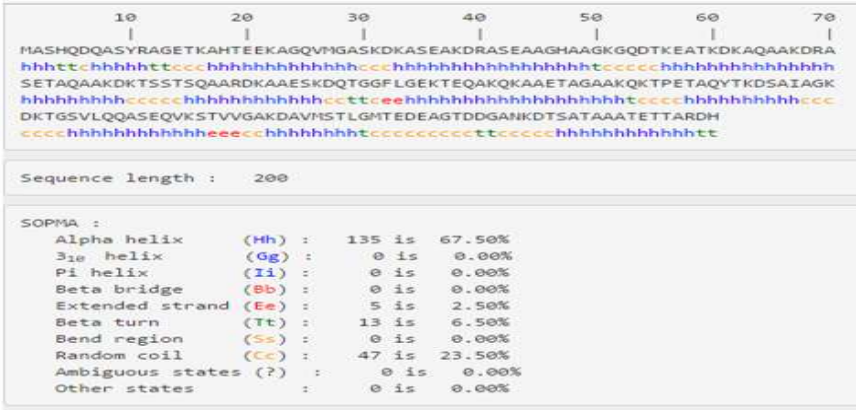

Figure 5a: Secondary Structure of LEA

Figure 5b: Structural Conformations of DREB

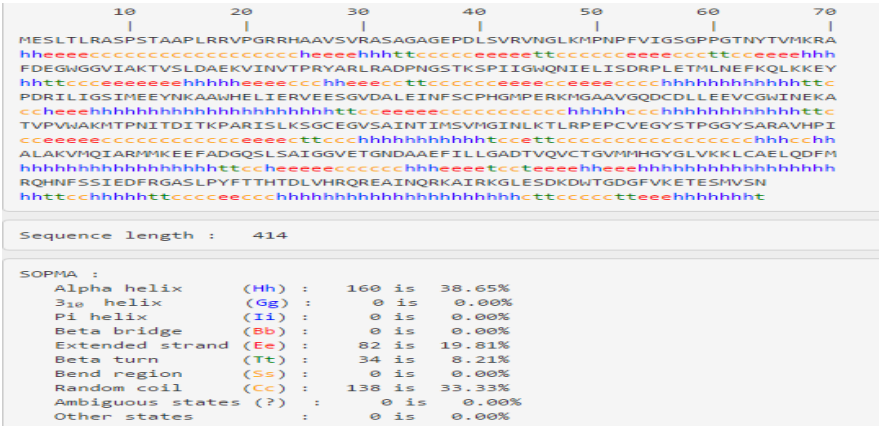

Figure 5c: Secondary Structural Characterization of OsDHODH1

Inference: All the above three proteins' secondary structure results indicate the conformational changes of the amino acids within the protein. The detailed composition of helix, sheets and random coils has been provided.

Tertiary Structure Prediction Using Phyre

Phyre compare the user entered sequences with the list of sequences from the PDB data bank in order to obtain the most suitable structure matching the sequence entered.

In addition to the results of the phyre tool displaying the best PDB id's the structures of the proteins can be studies using RASMOL visualization software.

The results of the Phyre and PDB indicate that the structure data related to the three proteins is not available in the database. Hence Homology modelling can be employed to develop the structures of the three gene products.

CONCLUSIONS

In the current work, a detailed analysis of the three proteins LEA, DREB and OsDHODH1 has been made insilico. The analysis was also used to compare the similarities if any existing among the three. The results show that the three sequences did not share any considerable similarity. The structural study was also performed to analyze the physicochemical characteristics of the proteins followed by the identification of their secondary structural conformations wherein all three proteins were found to be richer in helix than strands. The tertiary structure analysis revealed that the structures of the three proteins are not known and their prediction has to be made. The further modelling approach can be employed to develop the structures of the three proteins.

REFERENCES

1. *Rohit Joshi etal, "Transcription Factors and Plants Response to Drought Stress: Current Understanding and Future Directions", Frontiers in Plant Sciences, July 2016*
2. *Jianli Duan, Weiming Cai, "OsLEA3-2, an Abiotic Stress Induced Gene of Rice Plays a Key Role in Salt and Drought Tolerance", PLOS, July 2016*
3. *Rafi Shaik and Wusirika Ramakrishna, "Bioinformatic Analysis of Epigenetic and MicroRNA Mediated Regulation of Drought Responsive Genes in Rice", PLOS 7, 2017*
4. *Adam N. Famoso etal, "Genetic Architecture of Aluminum Tolerance in Rice (Oryza sativa) Determined through Genome-*

Wide Association Analysis and QTL Mapping", PLOS, Aug 2016

5. *Mglinets AV etal, "[Comparative sequence analysis of the LEA gene fragment in Pinus sibirica du tour and Pinus pumila (Pallas) regel]", Genetika, Feb 2014*
6. *Gasteiger E., Hoogland C., Gattiker A., Duvaud S., Wilkins M. R., Appel R. D., BairochA.;* Protein Identification and Analysis Tools on the ExPASy Server;
7. *Letunic I, Doerks T, Bork P (January 2012). "SMART 7: recent updates to the protein domain annotation resource". Nucleic Acids Res. 40 (Database issue): D302–5. doi:10.1093/nar/gkr931. PMC 3245027 PMID 22053084.*
8. *Yu-Duan Liao etal, "*
9. Oryza sativa *protein phosphatase 1a (OsPP1a) involved in salt stress tolerance in transgenic rice", Molecular Breeding, March 2016*